



Method development and application of object detection and classification to Quaternary fossil pollen sequences

Robin von Allmen^a, Sandra O. Brugger^{a,b,*}, Kai D. Schleicher^c, Fabian Rey^a, Erika Gobet^{d,e}, Colin J. Courtney Mustaphi^a, Willy Tinner^{d,e}, Oliver Heiri^a

^a Geoecology, Department of Environmental Sciences, University of Basel, Switzerland

^b Paul Scherrer Institute, Villigen, Switzerland

^c Biozentrum, University of Basel, Switzerland

^d Institute of Plant Sciences, University of Bern, Switzerland

^e Oeschger Centre for Climate Change, University of Bern, Switzerland

ARTICLE INFO

Handling editor: Dr Yan Zhao

Keywords:

Automated imaging
Convolutional neural networks (CNNs)
Fossil pollen
Lake sediment analyses
Late glacial pollen record
Pollen analysis
Quaternary vegetation reconstruction

ABSTRACT

The automation of fossil pollen analysis promises many advantages in handling large numbers of samples with less resource allocation. However, automation is often obstructed by the high abundance of organic and mineralogenic non-pollen debris in fossil pollen samples. We used a Convolutional Neural Network-based approach to detect pollen-like objects in digital images of prepared microscopic slides for fossil pollen analysis and subsequently classified them into nine pollen classes and the marker spore *Lycopodium*. We trained the object detection and the classification model independently with a newly developed dataset of annotated images of fossil pollen grains. The object detection model achieved average recall rates of 89.8 % and 75.5 % for pollen classes and *Lycopodium*, respectively. The classification model correctly categorizes fossil pollen images with >95 % accuracy. We applied the assembled pipeline to Late Glacial pollen samples using class-dependent thresholds to discriminate true pollen from non-pollen objects and compared automated count data for nine pollen types with manual pollen counts. For the selected pollen types, our results demonstrate the feasibility to replicate major fossil pollen changes with automated counts, even when the automated pipeline was applied to pollen samples from a different site than used to train the models. High correlations ($r = 0.97$) between the first two axes of Principal Component Analyses (PCA) calculated based on automated and manual counts and high correlation ($r = 0.93$) indicated by a Procrustes rotation analysis of the PCA results demonstrate that the two procedures reconstructed similar pollen patterns. While our automated approach is not yet able to achieve the taxonomic resolution of manual counts by expert analysts and is limited to selected pollen types, it provides a “proof of principle” that automated analyses can be applied to complex fossil pollen samples and to develop downcore stratigraphies. Automated analyses may with time lead to reliable pollen records. For instance, our pipeline can be further improved by adding more pollen classes, increasing the dataset of annotated images of fossil pollen grains, expanding the training data to rare pollen types, refining taxonomic resolution (e.g., separation of *Betula nana*-type or *Pinus*-types), and incorporating more challenging pollen types (e.g., *Juniperus*), to expand its application beyond reconstructing temporal changes in a few selected pollen types.

1. Introduction

Pollen analysis on natural archives such as lake and marine sediments, peat records, and ice cores is the most widely used palaeoecological approach to reconstruct Quaternary vegetation change, but may indirectly also provide information on landscape development,

early human impact, and climatic changes (e.g., Brechbühl et al., 2023; Brugger et al., 2019b; Dendievel et al., 2022; Felde and Birks, 2019; Galka et al., 2022; Githumbi et al., 2018; Rey et al., 2017). To date, pollen, spores, and other microfossils from lake sediments and other natural archives are manually analyzed by a human analyst through optical microscopy (Beug, 2004; Lang et al., 2023). The analyses require

* Corresponding author. Department of Environmental Sciences, University of Basel, Switzerland.

E-mail address: sandra.bruegger@unibas.ch (S.O. Brugger).

<https://doi.org/10.1016/j.quascirev.2024.108521>

Received 4 November 2023; Received in revised form 17 January 2024; Accepted 23 January 2024

0277-3791/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

a high level of palynological expertise, are time consuming, and need adaptation due to the variable quality of prepared microscope slides (Stillman and Flenley, 1996; Holt et al., 2011; Khanzhina et al., 2018). These tedious shortcomings may potentially become critical when aiming for contiguous high-resolution records with hundreds of samples (e.g., Camuera et al., 2019; Rey et al., 2019; Torres et al., 2013). In principle, automated analysis of fossil pollen records could be faster, allow potentially for more samples to be analyzed with higher pollen count sums, and may provide supplementary information on pollen morphology (Bourel et al., 2020). In addition, automated analysis would require less pollen expertise and may provide better consistency within a record and between records compared to manual pollen counts (Stillman and Flenley, 1996).

Several image-based automated techniques using geometric and texture features of pollen grains, and non-image-based pollen fluorescence techniques have been previously tested for pollen grain identification (Dell'Anna et al., 2009; Dunker et al., 2021; Holt and Bebbington, 2014; Holt and Bennett, 2014; Ivleva et al., 2005; Mitsumoto et al., 2010; Sevillano et al., 2020). First automated applications to modern pollen assemblages in other fields have shown promising results (e.g., Daunys et al., 2022; Lagerstrom et al., 2015; Sevillano and Aznarte, 2018). Such samples from honey or airborne pollen assemblages are generally characterized by excellent pollen preservation and often contain low taxonomic diversity and low amounts of non-pollen debris (Lagerstrom et al., 2015). To date, up to 47 pollen types have been distinguished in such modern samples with 98 % accuracy (Sevillano et al., 2020). However, most methodologies failed for fossil pollen assemblages from lake sediments due to the large floristic diversity, fragmented, folded, and clumped grains, and the issue of discriminating microfossils of interest from other debris in fossil samples (e.g., amorphous organic matter, algal or invertebrate remains; Holt et al., 2011; Holt and Bennett, 2014; Khanzhina et al., 2018). Therefore, applications to fossil pollen samples, such as those from lake sediments (Bourel et al., 2020; Milošević et al., 2020; Theuerkauf et al., 2023) or ice cores (Maffezzoli et al., 2023) are still in early stages (Polling, 2021).

Fully automated identification of fossil pollen assemblages from microscope images would require two steps. First, objects have to be detected on images (i.e., recognition of individual pollen-like objects within larger digital images or during automated screening of pollen slides). Second, the detected objects have to be classified (i.e., assigning taxonomic categories to individual images of pollen-like objects). Aside from the automated Classifynder (Holt et al., 2011; Lagerstrom et al., 2015), BeE-nose (El Azari et al., 2023) and BAA500 (Oteros et al., 2020) systems, most studies have focused on either classification (Astolfi et al., 2020; Bourel et al., 2020; Mahmood et al., 2023; Sevillano et al., 2020) or object detection (Johnsrud et al., 2013; Yamazaki et al., 2023) but have not truly automated the pollen counting methodology. Moreover, all the systems mentioned above have either no applicability in fossil pollen assemblages (e.g., El Azari et al., 2023; Holt and Bennett, 2014; Oteros et al., 2020) or have not yet been tested in this area. A recent, promising study (Theuerkauf et al., 2023) has developed automated image detection and classification methods based on individual fossil pollen samples, but not applied these to develop a fully automatically counted downcore stratigraphy of pollen counts that can be compared and evaluated relative to manually counted pollen assemblage data.

Here, we aim to advance automating fossil pollen counting by implementing (1) object detection of pollen in fossil sediment assemblages, (2) classification of fossil pollen grains using Convolutional Neural Networks (CNNs), and (3) combining the two steps in a pipeline to develop, to our knowledge, the first automatically analyzed pollen stratigraphy. We compare the results with manual pollen counts and discuss strengths and weaknesses of the approach as well as future developments that are necessary to further advance automated pollen detection in fossil pollen assemblages based on our approach.

2. Material and methods

The methodological workflow in this study consists of preparation of pollen slides from Late Quaternary sediment records from temperate European lowland lakes suitable for automated imaging; the development of an automated image acquisition profile, image preprocessing, object detection and classification that can be combined to a fully automated pipeline; and the development of tests, cross-validation procedures and applications to assess the performance of individual steps of the automated pipeline when applied to develop pollen count data from Late Quaternary lake sediments.

2.1. Study sites and preparation of pollen slides for automated imaging

We reused pollen samples and data published from the site Burgäschisee on the Western Swiss Plateau covering 15,700 to 10,900 calibrated ^{14}C years before present (cal yr BP) as a main record for developing and applying our methodology (Rey et al., 2017). The transition from Late Glacial tundra to boreal conifer forests at the beginning of the Bølling was studied in many pollen records in temperate Europe and is characterized by a few dominant pollen taxa with clear changes in the pollen stratigraphy (e.g., Rey et al., 2017; Lotter, 1999; Ammann et al., 2013) providing a suitable record with clear shifts in vegetation to test automated detection and classification approaches.

To train and evaluate our automated pipeline we prepared slides from 26 of the original sediment samples of the Burgäschisee sequence, that were originally prepared using standard pollen sample techniques and mounted in glycerol (Rey et al., 2017). For samples with high pollen densities, we prepared one slide with equal amounts of images for training and application. In cases of moderate pollen densities, two separate microscopic slides were prepared (one for training, one for application), while low-density samples (less than 50 pollen grains per slide) required three slides (one for training and two for application). The slides had to fulfill the following criteria to be suitable for imaging purposes: pollen grains have to be visible within the same focal plane and should not be obstructed by non-pollen debris, no overlap or clustering of pollen grains should be apparent, per core depth at least 50 pollen grains should be detectable, and the samples should be sufficiently stained to clearly see pollen features. Samples were diluted using glycerol and stained with safranin until the criteria above were met for each depth unit.

In addition to the Burgäschisee site, we applied the methodology to the Hüttwilersee record, a sediment sequence from the Eastern Swiss Plateau (~100 km from Burgäschisee; Wick et al., 2024) to test the approach on a completely independent pollen record. We prepared three samples from the Bølling that are characterized by high amount of *Betula* pollen and three samples from the Younger Dryas with abundant *Pinus* pollen (Wick et al., 2024). The microscopic slide preparation followed the criteria described above.

2.2. Architecture of the automated pipeline

The automated pipeline for pollen detection and classification assembled in this study consists of i) image acquisition and preprocessing, ii) an object detection model followed by iii) a CNN for classification, thresholding steps after object detection and classification, and iv) an output layer with merged data (see Fig. 1).

i) Image acquisition and preprocessing: To achieve true automation in fossil pollen analyses, a critical advancement entails the utilization of automated imaging rather than manually capturing individual pollen images (Khanzhina et al., 2018). We used a Zeiss AxioScan Z.1 slide scanner, an automated widefield microscope that can sequentially process up to one hundred slides. The slides were scanned using a Plan-Apochromat 20x objective with a numerical aperture of 0.8 and a 3CCD Hitachi HV-F202SCL color camera. Each scanned slide consisted

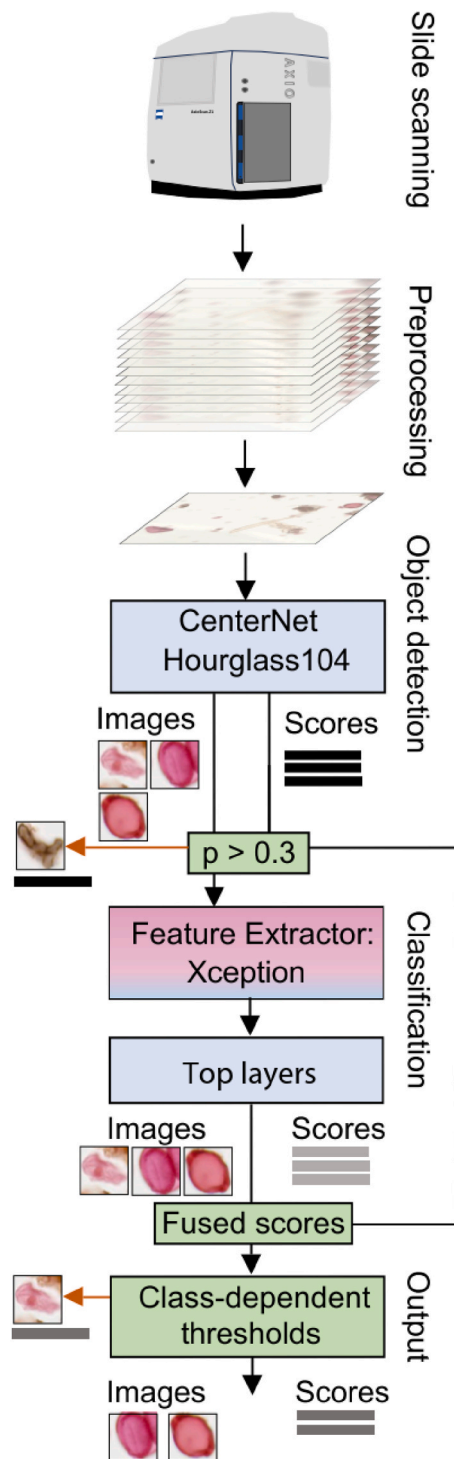


Fig. 1. Architecture of the automated pipeline. Blue colored layers determine layers in the object detection and classification models that contained trainable parameters. Pink colored layers were fixed in the feature extractor during Transfer Learning. Green colored boxes show positions for thresholds to filter non-pollen debris and pollen of poor quality after object detection and during fusion of the models.

of ~6500 individual rectangular image tiles with dimensions of 1600 pixels in width and 1200 pixels in length. We retiled the rectangular image tiles to squared images with dimensions of 1200 by 1200 pixels using the pylibCZlrw (version 3.3.0, Carl Zeiss AG), python-bioformats (version 4.0.7, Linkert et al., 2010) and the cztile (version 0.0.2; Carl Zeiss AG) packages in python. Each squared image tile consisted of

multiple Z-planes (12 planes in intervals of 2.8 μm). We compressed the z-planes to a single image using a focus stacking algorithm modified from Matsuo et al. (2022). This process conceptually consisted of three steps: first, partitioning the three-dimensional image stack into smaller subregions (120 by 120 pixels) across the x- and y-axis to better capture the three-dimensional surface of small pollen grains, second, applying the Sobel-Filter to identify the sharpest Z-plane per subregion of the image stack, and third, overlaying the subregions using a maximum intensity projection. **ii) Object Detection:** This step locates objects of interest within the larger input image and asserts class labels to the detected objects. After observing low classification precisions in initial trials, we were only interested in maximizing recall, which is the proportion of detected relevant objects relative to the sum of detected and missed objects. CenterNet Hourglass104, an object detection model, has reportedly performed well on the standard image dataset COCO (Common Objects in Context dataset; Lin et al., 2014) and overall performed best in our initial trials. The CenterNet Hourglass104 was trained to perform detection on nine important pollen taxa (Apiaceae, *Artemisia*, *Betula*, *Amaranthaceae*, *Cyperaceae*, *Helianthemum*, *Pinus*, *Poaceae*, *Salix*) observed in Late Glacial European pollen sequences and the marker spore *Lycopodium clavatum*, used in standard pollen preparation protocols (Brugger et al., 2019a; Lang et al., 2023). We split the three pollen taxa *Artemisia*, *Betula*, and *Poaceae* into separate classes for equatorial and polar view, as initial experiments indicated that pollen grains orientated in equatorial view were often misclassified or had low confidence levels in models trained with mixed view classes (see Supplementary Fig. S1 for complete list of the 10 pollen classes and criteria). Images with detected pollen-like objects were run through a confidence threshold filter to avoid non-pollen objects. We prioritized the value for this threshold to capture all true pollen grains with suitable image quality in order to optimize recall. Bounding boxes of objects with confidence scores above the threshold (confidence value > 30 %) were transformed from rectangular to squared dimensions to prevent distortions further down in the pipeline. The squared box coordinates were used to crop out the relevant parts in the image and the presumed pollen object was then classified by the classification model. **iii) Classification:** The classification model will by default classify each object to one of the training classes. Any non-pollen palynomorph (NPP) detected as pollen-like objects by the object detection model in the first step will therefore be assigned to the most likely pollen class. Therefore, we added four NPP classes (NPP Type 1–4) to the classification model that were frequently detected by the object detection model (Supplementary Fig. S1.). We chose the Xception (Chollet, 2017) model for classification following previous studies (Olsson et al., 2021; Mahmood et al., 2023). We used the convolutional layers as feature extractor and built a network of dense layers on top of it for classification (Fig. 1). The Dropout Layer in between the dense layers mitigates overfitting to training images (Zeng et al., 2023). We used pretrained weights in the keras package (version 2.6.0; Kim et al., 2022) for the feature extractor. This process of the so-called Transfer Learning approach is useful when the size of the training dataset is limited (Pan and Yang, 2010). **iv) Merging Output:** The output from the classification model can be used to estimate pollen numbers per class by using the highest confidence values as class label for each pollen-like object. These confidence values are based on the combined output of the classification and the object detection model by averaging the confidence values per class. We compared the performance of one global threshold (configuration called fused global threshold) and class-dependent thresholds (configuration called fused class-dependent thresholds) determined by examining the confidence values of pollen images in the validation dataset (Supplementary Fig. S3; Pardasani and Awasthi, 2020). These thresholds allow to omit false positive predictions by the model (e.g., debris that has similarities with pollen) and only regard predictions exceeding the threshold to be counted as true positives.

2.3. Training of individual modules in the pipeline

Object Detection: All training slides together yielded 192,934 individual square images that were manually annotated by two palynologists using the open-source image annotation tool LabelImg (Tzutalin, 2023). Annotation resulted in 2326 annotated images with pollen grains fulfilling the criteria in Table 1. A balanced dataset with equal numbers of training images for each class is preferred because class imbalances can cause class-specific overfitting to overrepresented classes (Japkowicz and Stephen, 2002; He and Garcia, 2009). We used 35 images per class, which has shown to be sufficient for training purposes (Astolfi et al., 2020; Li, 2004; Sevillano and Aznarte, 2018). A set of augmentations was performed to create a mixed dataset of original and artificial pollen grain images for taxa with insufficient numbers of annotations (i. e., *Helianthemum* and *Amaranthaceae*) following Mahmood et al. (2023).

20 % of all images per class were put aside for validation purposes resulting in 504 images for training and 126 images for validation. We

Table 1

List of criteria for the pollen classes to establish an image collection for training the object detection and classification models. Note that pollen classes for *Artemisia*, *Betula*, and *Poaceae* were further divided into subclasses for respective polar and equatorial view to facilitate grain recognition. t. = type.

Pollen class and corresponding pollen types following Beug (2004)	Criteria for picture inclusion
Apiaceae: <i>Anthriscus sylvestris</i> -t., <i>Astrantia</i> -type, <i>Berula erecta</i> , <i>Bupleurum</i> -t., <i>Carum carvi</i> , <i>Chaerophyllum</i> -t., <i>Daucus</i> , <i>Eryngium</i> , <i>Falcaria</i> -t., <i>Heracleum</i> , <i>Laserpitium latifolium</i> -t., <i>Ligusticum mutellina</i> , <i>Peucedanum</i> -t., <i>Pimpinella major</i> -t., <i>Seseli</i> -t., <i>Sison amomum</i> , <i>Trinia glauca</i> , <i>Apiaceae</i> indet, <i>Apium</i>	• Colporate aperture clearly visible • Full grain, not broken • Sharp focus, no other pollen visible
Artemisia: <i>Artemisia</i>	• Full grain, not broken • Sharp focus, no other pollen visible • Polar view: Wall thinning towards colpus clearly visible for all sides • Equatorial view: Wall thinning towards colpus clearly visible
Betula: <i>Betula alba</i> -t., <i>Betula nana</i> -t.	• Full grain, not broken • Sharp focus, no other pollen visible • Polar view: All 3 pori visible, clear vestibulum • Equatorial view: 1–2 pori clearly visible
Amaranthaceae: <i>Amaranthaceae</i>	• Pori clearly visible • Full grain, not broken • Sharp focus, no other pollen visible
Cyperaceae: <i>Cyperaceae</i>	• Pores clearly visible • Characteristic grain shape visible, not folded • Full grain, not broken • Sharp focus, no other pollen visible
Helianthemum: <i>Helianthemum</i>	• Minimum one colpore clearly visible • Full grain, not broken • Sharp focus, no other pollen visible
Pinus: <i>Pinus sylvestris</i> -t., <i>Pinus cembra</i> -t.	• Full saccus, not broken • clearly visible ornamentation on wing • Sharp focus, no other pollen visible
Poaceae: <i>Poaceae</i> , <i>Cereal</i> -t., <i>Hordeum</i> -t., <i>Triticum</i> -t., <i>Avena</i> -t.	• Full grain, not broken • Sharp focus, no other pollen visible • Polar view: Pori in top view • Equatorial view: Pori in side view
Salix: <i>Salix</i>	• Prolate shape, colpi visible • Reticulum with margo visible • Full grain, not broken • Sharp focus, no other pollen visible
Exotic marker: <i>Lycopodium clavatum</i>	• Clearly defined surface ornamentation or clearly defined columellae • Full grain, not broken • Sharp focus, no other pollen visible

used further data augmentation to increase the number of images for training and validation (Bourel et al., 2020; Mahmood et al., 2023). We employed Gaussian blur and noise, random saturation adjustment, haze reduction and image sharpening (Mahmood et al., 2023; Supplementary Fig. S2). Additionally, every image was randomly flipped and rotated. We augmented the data once and assured the quality of the images again (Mahmood et al., 2023). Finally, the training dataset consisted of 3020 augmented images and the validation dataset of 756 images.

We trained the detection model with a jupyter notebook from github (Kamal et al., 2023) that employs the tensorflow object detection package (version 0.1; Srivastava et al., 2022). Since the training of the Hourglass104 model (Duan et al., 2019) needs considerable computing capacity, we deployed training to Paperspace, a cloud-based computing website, on a Nvidia RTX A600 machine (Jermain et al., 2016; Mathai-Davis and Trudeau, 2019). The model was trained for 10,000 epochs with an initial learning rate of 0.001 and manual learning rate declining to 0.0001 after 3500 epochs and to 0.00001 after 5000 epochs. We restricted the maximal number of detected boxes to 100. For details the reader is referred to the code on GitHub (<https://github.com/RobinVonAllmen/FOSSILPOLLEN>).

Classification: We obtained cropped pollen grain images for training and validation using the annotations in the original object detection dataset. We used the same data augmentation techniques described above. We employed 5-fold cross-validation by splitting the dataset into five subsets (doi: 10.5281/zenodo.8398543) and using each part as validation set in a separate training loop following Mahmood et al. (2023). The model was trained on a NVIDIA GeForce RTX 3050 Ti laptop GPU using the tensorflow (version 2.8.0; Abadi et al., 2016) library in VisualStudio Code (version 1.79.2; Clow, 2018). We used the optimized parameters input size (140,140,3), initial learning rate equal to 0.00005 and applied an optimized stochastic gradient with the Adam algorithm (Kingma and Ba, 2014) after training and validation of the classification model. For further details the reader is referred to the code on GitHub (<https://github.com/RobinVonAllmen/FOSSILPOLLEN>).

2.4. Evaluation procedure

We first evaluated the classification model performance since it is the key step in the pipeline architecture to filter out and correctly classify input images. Next, we evaluated whether the object detection task is effective in extracting the majority of pollen grains from the input images and that the relative abundances are comparable with manual values from Rey et al. (2017). Lastly, we examined the pipeline's potential use for developing reliable pollen records.

Evaluation of Classification Model: We assembled a test image dataset for classification by retrieving a set of pollen images among the imaged slides from Burgäschisee (Rey et al., 2017) that is used for testing purposes only (i.e., the images were not used for training the model). We evaluated the performance on 20 images per class fulfilling the criteria in Table 1 and on a mixed dataset consisting of the images from the first experiment but also 20 images per pollen class with visible characteristics and low image quality (Bourel et al., 2020). For visual confirmation of training success, we employed the Style-Gam algorithm and activation visualizations of different convolution layers (Chollet, 2020; Fu et al., 2020).

Evaluation of Object Detection: To evaluate whether our model detects all categorized pollen grains across our samples, we analyzed the predictions for 10 % of all images in every 4th sample depth of our test dataset. We report the performance in Recall and Precision. Recall is the ratio between correctly detected pollen and the sum of all pollen in the images. We define precision as the ratio between correctly identified pollen and the sum of all as pollen classified objects. We evaluated the detection of *Lycopodium* with the same procedure.

Comparison with manual count data: We plotted the relative abundances of the Burgäschisee original counts (Rey et al., 2017) against the output of the automated pipeline in a pollen stratigraphy. This visual

evaluation aimed to determine if the automated pipeline’s results match manual counts. For both, the manual and the automated counts, we calculated pollen percentages using the total pollen sum of the nine pollen classes and 95 % confidence intervals following Maher (1972) and Maher et al. (2012). *Lycopodium* spores were excluded from the pollen sum, but the percentages were calculated relative to the pollen sum for evaluation. The same procedure as above was used on the samples from Hüttwilersee (Wick et al., 2024). We applied Principal Component Analysis (PCA) to the Burgäschisee count data to compare the similarity of the automated pollen counts with manual pollen counts from the same sample depth. The PCA was calculated for the manual-counts-based percentage abundances of pollen counts and the automated count data were passively plotted in this ordination (Birks et al., 2012). In this analysis, the distance between samples based on manual and automated pollen counting in a site score scatter plot of the PCA represents an estimate of the dissimilarity between the different pollen assemblages. PCA was also used to examine whether pollen counts based on the automated pipeline were able to capture the underlying data structure represented in the manually counted pollen data. For this purpose, two separate PCAs of the Burgäschisee count data were calculated, the first based on the manually counted pollen samples, the second based on the counts from the automated procedure. We evaluated the correlation of corresponding sample scores for PCA axis 1 and 2 between the automated and the manual dataset using Pearson correlation coefficients (Birks et al., 2012). We further investigated the similarity between the PCAs with a symmetric Procrustes rotation and used PROTEST (9,999,999 permutations) to test the statistical significance of the similarity (Peres-Neto and Jackson, 2001). We expected a high correlation between the PCA axes of the two analyses and in the Procrustes rotation analysis if the two pollen counting procedures (manual, automated) capture the same underlying paleoecological signal in the Burgäschisee record. PCAs were calculated based on covariance matrices and square-root transformed percentage data of identified pollen types in the program CANOCO for Windows (Ter Braak and Smilauer, 1997-2019). For Procrustes rotation analysis PCAs were recalculated using the same options in the R environment using the package *vegan* (Oksanen et al., 2022).

3. Results

Visual examination revealed that the imaging process yielded images with clearly visible distinguishable characteristics for most pollen classes (Supplementary Fig. S5). The colpi/apertures of *Helianthemum* grains were difficult to capture in focus due to the large grain size and rounded shape. We excluded *Juniperus* pollen, originally planned as a pollen class, from the dataset because the image resolution was not sufficient to distinguish the gemmate surface structure and the variable geometry of the pollen (Beug, 2004). Furthermore, we excluded *Populus*, a taxon frequent in some European Late Glacial sequences (Ammann

et al., 2013) and dominant in North American records, because poplar was not one of the main taxa in the Burgäschisee record of Rey et al. (2017).

3.1. Object detection model results

On the test data images, the trained object detection model performed with an average recall of 89.8 % (range 87.6 %–94.7 %) for pollen and 75.5 % (range 66.7 %–82.4 %) for *Lycopodium* spores, respectively (Table 2). Recall rates indicate that the model missed on average 10.2 % of all pollen grains and 24.5 % of the *Lycopodium* spores on the scanned images. Precision varied between 41.9 % and 62.6 % (average = 61.6 %) for pollen detection (Table 2). We found no obvious bias in recall and precision rates between the different pollen classes. For *Lycopodium*, the precision was higher but more variable between samples (average = 88.4 %, range = 73.7 %–100 %).

3.2. Classification model results

After Transfer Learning the Xception model reached an initial training and validation accuracy of 98.5 % and 96 %, respectively. With further fine-tuning by including the last hundred layers of the feature extractor into the training loop following Subramanian et al. (2022), training increased to 100 % and validation to 99.5 % (Fig. 2). Output visualizations with the Style-GAM algorithm (Fig. 3) and activation visualizations (Supplementary Fig. S1) show that the model trained class-specific morphologies e.g., the porate apertures in the *Amaranthaceae* class and the surface structure of *Pinus* sacci.

During application to the test dataset with good quality images the trained model reached an average accuracy of 97 % across all 5 cross-validation cycles. The first cross-validation cycle (C1) resulted in a per-class accuracy of 100 % for all classes, which equals to no misclassifications (Fig. 4). The other cross-validation cycles (C2–C5) resulted in one to three misclassifications each. One *Amaranthaceae* was misclassified and assigned to *Poaceae*, one *Salix* was assigned to *Pinus* and two *Helianthemum* were misclassified as *Salix*.

The first cross-validation cycle (C1) also performed best on mixed quality images with an overall accuracy of 98.8 % (Fig. 5). With mixed quality images, the model classified all images in six pollen classes and the *Lycopodium* class correctly. For *Amaranthaceae* and *Pinus* one image was misclassified to *Poaceae* and *Lycopodium*, respectively and for *Helianthemum* one image was assigned to *Betula* and one to *Salix*. For the other cycles overall accuracy was also very high (>95 %; Fig. 5). Overall, the model occasionally misclassified *Amaranthaceae* as *Poaceae*, *Betula* as *Poaceae* and *Helianthemum* as *Artemisa*. All other misclassifications seemed sporadic and not clustered for a particular taxon (Fig. 5).

Table 2
Object Detection evaluation through manual validation of predictions. The Recall rates represent the proportion of detected objects of interest (true positives) relative to the sum of detected objects of interest and undetected objects (false negatives) for pollen and *Lycopodium* marker (Lyc.). Precision rates represent the proportion of detected objects of interest (true positives) relative to the sum of all detected objects in this category (true positives and false positives) for pollen classes and *Lycopodium* marker.

Sample age [cal yr BP] ^a	True positives pollen/Lyc. [counts]	False negatives pollen/Lyc. [counts]	False positives pollen/Lyc. [counts]	Recall pollen/Lyc. [%]	Precision pollen/Lyc. [%]
12,027	378/70	42/15	246/8	90.0/82.4	60.6/89.7
12,730	263/28	22/11	157/0	92.3/71.8	62.6/100
13,406	381/14	54/7	122/5	87.6/66.7	50.0/73.7
14,018	54/23	5/8	72/5	91.5/74.2	42.9/82.1
14,645	36/35	2/16	50/4	94.7/68.6	41.9/89.7
15,523	32/52	5/15	66/7	86.5/77.6	50.0/88.1
Sum of all samples	1144/222	130/72	713/29	89.8/75.5	61.6/88.4

^a Chronology of Burgäschisee samples following Rey et al. (2017).

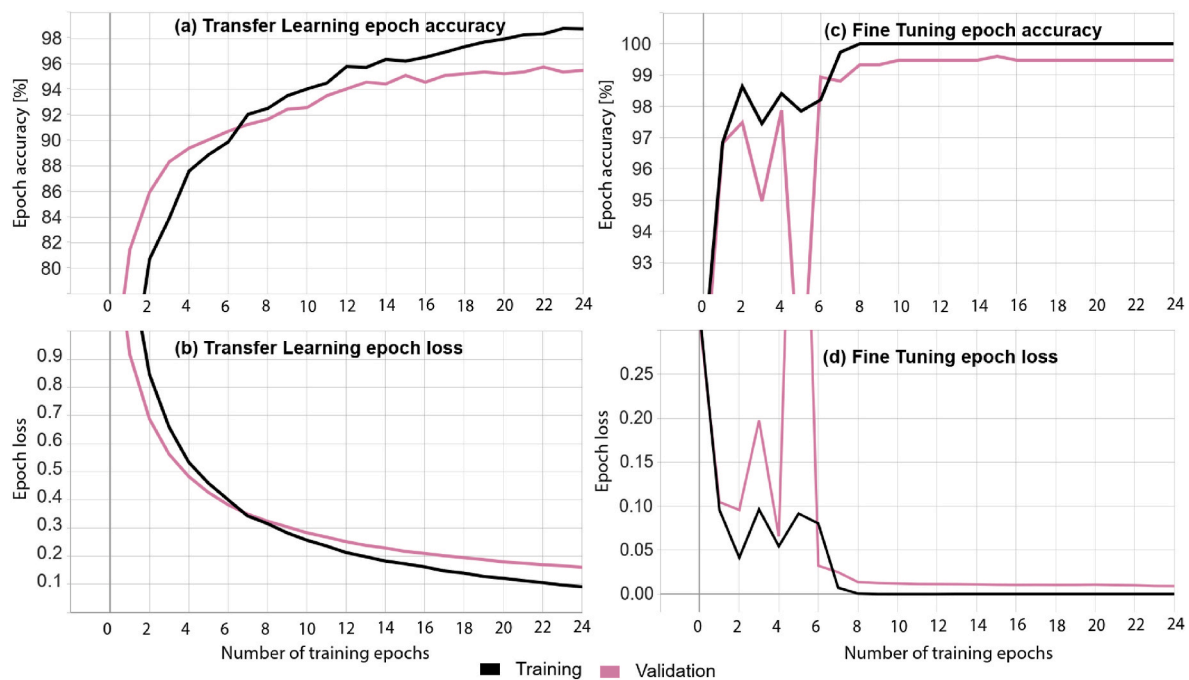


Fig. 2. Training progress of Classification with Xception (Chollet, 2017). Changes in accuracy (a) and loss (b) while training the fully-connected layers only (Transfer Learning). Progression of accuracy (c) and loss (d) during fine-tuning by including the last hundred layers of the feature extractor in the training loop following Subramanian et al. (2022). Accuracy refers to the ratio between correct predictions and the total number of predictions in the training data while loss refers to the error between the predicted and true value. Epochs refer to the number of iterations during training and validation.

3.3. Application of the automated pipeline to the Burgäschisee record

Lycopodium percentages generated from the individual parts of the pipeline (i.e., object detection, classification and fused global threshold) as well as the final configuration of the pipeline using fused scores with class-dependent thresholds (fused class-dependent thresholds) closely match the manual *Lycopodium* count data with percentages ranging between <5 % and ~80 % along the record (Supplementary Fig. S4).

All four tested configurations show similar temporal trends of pollen classes during the Late Glacial at Burgäschisee, with generally low percentages of all herbaceous classes and *Salix*, as well as high abundances of *Betula* and *Pinus* class (Supplementary Fig. S4). However, the configuration with fused class-dependent thresholds predicts the pollen percentages best among the tested configurations in comparison with the manual pollen count data by Rey et al. (2017).

The pipeline with class-dependent thresholds characterizes the pollen assemblages in the oldest samples at Burgäschisee covering the Oldest Dryas period and the earliest part of the Bølling with ~25 % *Betula*, up to 15 % Poaceae, and low amounts of *Artemisia*, Cyperaceae, and Amaranthaceae (Fig. 6). The automated counts closely match the manual count data of Rey et al. (2017), except for the oldest sample, where *Helianthemum* class is estimated ~40 % while manual count data contained 4 % *Helianthemum* (Fig. 6). Likewise, Amaranthaceae (~10 %) and *Salix* (peak at 20 % vs. <10 %) are overrepresented in the automated dataset compared to manual counts (<5 %). In contrast, Cyperaceae percentages are predicted to be considerably lower (3 %) compared to the manual count data (~18 %) and *Artemisia* is underrepresented.

Automated counts suggest around ~30 % *Betula* class, which closely match manual counts (Fig. 6). Poaceae increases to ~30 % in both automated and manual count data in samples dated to ~14,600 and ~14,200 cal yr BP. Furthermore, the automated counts suggest highest amounts of Apiaceae class of the record (~4 %) when Apiaceae percentages are highest in the manually counted record (i.e., ~2 %).

Samples in the Bølling *Betula* phase are characterized by a rapid increase of *Betula* class percentages to >70 % in automated and manual count data, which are subsequently replaced by increasing amounts of

the *Pinus* class towards the Younger Dryas samples (Fig. 6). Although, the automated pipeline predicts the trends in the pollen assemblages correctly, the change occurs more rapidly in the automated count data, resulting in several samples of the Allerød dated to ~13,600 to ~12,800 cal yr BP, in which the *Betula* class is underrepresented in the automated counts (up to ~45 % higher in the manual count data; Fig. 6). Likewise, the *Pinus* class with values > 80 % is overrepresented in the automated count data during the Allerød when *Pinus* increases only slowly in the manual count data to reach similar percentages as in the automated counts at the onset of the Younger Dryas. Samples from the Younger Dryas and Early Holocene (after ~11,800 cal yr BP) are generally characterized by high abundances of *Pinus* class pollen in the automated count data, which remains >75 % when the manual count data suggests a decrease to ~30 % *Pinus* (Fig. 6). Simultaneously, *Betula* pollen in the manual count data increases to ~50 % in the youngest samples, an increase that is not predicted to the same extent by the automated counts of the class-dependent threshold model. Herbaceous pollen classes and *Salix* remain at overall low percentages in the automated dataset in agreement with the manual count data.

The PCA for manual count data of Rey et al. (2017) clearly separates samples from the Oldest Dryas and earliest Bølling from those of the Bølling and Allerød *Betula* phase, as well as from those of the Allerød *Pinus* phase, and the Younger Dryas-Holocene samples (Fig. 7A). The species scores (arrows) indicate that the samples group as expected relative to the main pollen classes such as herbaceous classes (e.g., Poaceae, *Artemisia*) grouped with *Salix* and distant from *Betula* and *Pinus*. The passively plotted sample scores of the automated counts show a similar pattern as the manually counted samples (Fig. 7A). Sample scores of the Oldest Dryas automated count data group closely with the Oldest Dryas manual samples scores and in the old part of the diagram with high species scores of the herbaceous pollen classes and *Salix*. Likewise, sample scores based on manual and automated counts of Allerød and Younger Dryas-Holocene samples group closely together along PCA axis 1 and are distant from the Oldest Dryas samples. The axis 2 gradient is less well pronounced in the automated samples but nevertheless clearly separates sample scores with high *Betula*

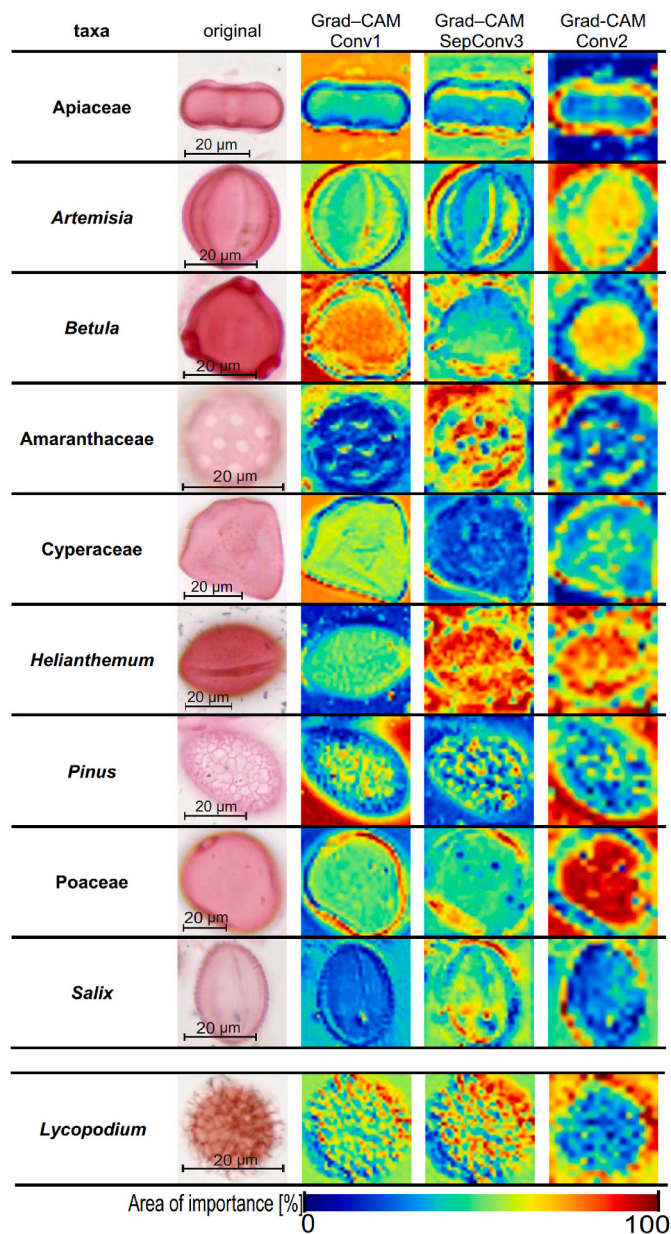


Fig. 3. Grad-Cam activation map (Chollet, 2020) on separable convolution layers in block 1 to 3 of the Xception model (Chollet, 2017). Colors indicate pixel-wise contribution to the classification output for the correct class in different layers as area of importance in percent (Mahmood et al., 2023). Three layers of representative images for each pollen class and the *Lycopodium* marker are presented.

percentages from the other samples. Furthermore, samples with high *Pinus* abundances, mainly from the Allerød and Younger Dryas, are also clearly separated from other samples along axis 2.

The axis 1 sample scores of the automated and manual count data are strongly and linearly correlated ($r = 0.97$) and confirm that component 1 in both datasets is similar (Fig. 7B). Axis 2 sample scores are also positively correlated but the correlation is weaker ($r = 0.86$) and there is a slightly higher variability in the relationship. The comparison of the PCAs based on automated and manual count data with Procrustes rotation likewise confirms a high similarity between the two analyses and datasets (p-value < 0.0000001 , correlation in a symmetric Procrustes rotation: 0.927). The results indicate that the records based on both, the manual and automated counts, capture the same underlying paleoecological signal associated with the major vegetation changes in the Late Glacial.

3.4. Application to independent samples from Hüttwilersee

We applied the automated pipeline (fused scores with class-dependent thresholds), trained on the pollen images from the Burgäschisee sequence, to pollen samples from the independent archive of Hüttwilersee to assess the performance on another Late Glacial sequence. The automated pipeline correctly assigned most pollen grains from the Bølling samples to the *Betula* class ($\sim 70\text{--}75\%$) and from the Younger Dryas samples to *Pinus* ($\sim 80\text{--}95\%$) (Fig. 8). Few pollen grains across all samples were assigned to herbaceous pollen classes and *Salix* ($< 5\%$ in all samples). These results are comparable to manual counts by Wick et al. (2024), where *Betula* pollen is abundant in Bølling samples and *Pinus* pollen dominates Younger Dryas pollen assemblages. However, *Pinus* counts were overestimated in the sample from 12,417 cal yr BP, where the model assigned 95 % of all grains to *Pinus*, while manual counts contained 70 % *Pinus*, and higher percentages of Poaceae and *Betula*. *Betula* was slightly overestimated in all three Bølling samples (5–10 % higher compared to manual counts) where it is the main component of the pollen assemblages.

The *Lycopodium* marker percentages were underestimated by the automated pipeline in all six samples (Fig. 8). In the Bølling samples the estimates ranged between 10 % and 14 % while values in the manual counts ranged from 20 % to 24 %. In the three samples from the Younger Dryas, *Lycopodium* was estimated up to four times lower compared to manual counts with values between 8 % and 14 % while manual counts contained 40–50 % *Lycopodium*.

4. Discussion

Most previous studies have focused either on object detection (e.g., Gallardo-Caballero et al., 2019; Johnsrud et al., 2013; Kubera et al., 2022; Yamazaki et al., 2023) or classification (e.g., Adaimé et al., 2023; Daunys et al., 2022; Polling, 2021; Sevillano et al., 2020). Studies that combined object detection and classification of pollen assemblages have not previously been successfully applied to series of fossil samples presumably due to the complex organic matrix (with many different remains) in pollen slides that make automatic counting more challenging compared to (clean) modern pollen assemblages (e.g., Bourel et al., 2020). A recent study has applied pollen detection and classification algorithms to individual fossil pollen samples (Theuerkauf et al., 2023), but without quantitatively comparing automatically generated with manually counted assemblage data as we do in our study. Our approach consists of a three-step pipeline which detects pollen-like objects in fossil samples and then classifies the detected pollen-like objects. The final configuration of our pipeline (fused scores with class-dependent thresholds) has the potential to detect changes in pollen compositions, to replicate fossil pollen signals in time series, and to complement manual counting efforts of Quaternary fossil pollen (Simpson and Birks, 2012). To our knowledge, this is the first study that shows the successful application of neural networks to develop Quaternary fossil pollen sequences across time intervals characterized by major vegetation changes (Holt et al., 2011).

4.1. Comparison of the individual pipeline modules with previous work

The trained object detection model in our study was effective in recovering pollen (recall rate of 89.8 %) and *Lycopodium* (recall rate 75.5 %), respectively (Table 2). These recall rates indicate that the model detected the vast majority of pollen correctly as pollen objects and are comparable with previous automated pollen detection studies that report recall rates between 40 % and 100 % (e.g., Astolfi et al., 2020; El Azari et al., 2023; D'Orazio, 2022; Gallardo-Caballero et al., 2019; Johnsrud et al., 2013; Kubera et al., 2022; Olsson et al., 2021; Oteros et al., 2015, 2020; Polling, 2021; Theuerkauf et al., 2023; Valiente et al., 2023; Yamazaki et al., 2023). The reported recall values in previous studies vary with number of taxa and the taxonomic

		Predicted class label										
		Apiaceae	Artemisia	Betula	Amaranth.	Cyp.	Heli.	Pinus	Poaceae	Salix	Lyc.	NPP
True class label	Apiaceae	C1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C2	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C5	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Artemisia	C1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C3	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C4	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C5	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Betula	C1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C4	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Amaranth.	C1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		C2	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		C3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		C4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		C5	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	Cyp.	C1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		C2	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		C3	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		C4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		C5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	Heli.	C1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
		C2	0.0	0.0	0.0	0.0	0.0	88.9	0.0	11.1	0.0	0.0
		C3	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
		C4	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
		C5	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	Pinus	C1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		C2	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		C3	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		C4	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		C5	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	Poaceae	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
		C2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
		C3	0.0	0.0	0.0	5.0	0.0	0.0	95.0	0.0	0.0	0.0
		C4	0.0	0.0	0.0	5.0	0.0	0.0	95.0	0.0	0.0	0.0
		C5	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	Salix	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		C2	0.0	0.0	0.0	0.0	0.0	5.0	0.0	95.0	0.0	0.0
		C3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
		C4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
		C5	0.0	0.0	0.0	0.0	0.0	5.0	0.0	95.0	0.0	0.0
	Lyc.	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		C2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		C3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		C4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		C5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0

Fig. 4. Confusion matrix comparing predicted and true values for good quality pollen images in percentages of all images per class. Five rows per predicted label across five cross-validation cycles are presented (C1–C5; Mahmood et al., 2023). The first split (purple) demonstrates the highest performing split. NPP = Non-pollen palynomorph.

resolution (e.g., Olsson et al., 2021; Oteros et al., 2020; Sevillano et al., 2020), the image resolution (e.g., Olsson et al., 2021; Valiente et al., 2023), and the amount of training data (e.g., El Azari et al., 2023; Yamazaki et al., 2023). However, average precision values of 61.6 % (Pollen) and 88.4 % (*Lycopodium*) indicate that our model also classified a considerable number of non-pollen objects as pollen-like objects. While very few pollen-like objects were missed (i.e., high recall rates), it also meant that further steps during classification (e.g., class-dependent thresholding) were necessary to eliminate non-pollen objects from the images detected as pollen.

Our approach faced inherent class-imbalances of fossil pollen assemblages for the training set (i.e., either *Betula* or *Pinus* were dominant in most samples) and morphologic variation due to taphonomy

processes over time (Pennington, 1996). Although, we have not observed differences in recall rates among the nine pollen classes, the marker class *Lycopodium* shows lower recall and higher precision rates in all samples. One reason for the differences in recall rates could be the class-imbalance between the nine pollen classes and a single marker class in the training dataset. Olsson et al. (2021) have observed a similar phenomenon that recall increases with several classes merged to one large class. Furthermore, *Lycopodium* spores and pollen grains have different morphological features that can influence predictive values. Some published pipelines incorporate a form of detection or localization of pollen-like objects that use techniques other than recall to assess model performance (e.g., Dell'Anna et al., 2009; Dunker et al., 2021; Maffezzoli et al., 2023; Mitsumoto et al., 2010; Pappas et al., 2003). For

		Predicted class label										
		Apiaceae	Artemisia	Betula	Amaranth.	Cyp.	Hel.	Pinus	Poaceae	Salix	Lyc.	NPP
Apiaceae	C1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C2	97.5	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0
	C3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C4	97.5	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0
	C5	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6
Artemisia	C1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C3	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C4	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C5	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Betula	C1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C2	0.0	0.0	95.0	2.5	0.0	0.0	0.0	2.5	0.0	0.0	0.0
	C3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C4	0.0	0.0	97.5	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0
	C5	0.0	0.0	97.5	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0
Amaranth.	C1	0.0	0.0	0.0	97.5	0.0	0.0	0.0	2.5	0.0	0.0	0.0
	C2	0.0	0.0	0.0	97.4	0.0	0.0	0.0	2.6	0.0	0.0	2.6
	C3	0.0	0.0	0.0	92.1	0.0	0.0	0.0	5.3	0.0	2.6	5.3
	C4	0.0	0.0	2.5	92.5	0.0	0.0	0.0	5.0	0.0	0.0	0.0
	C5	0.0	0.0	2.5	95.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0
Cyp.	C1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C2	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C3	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Heli.	C1	0.0	0.0	3.3	0.0	0.0	93.3	0.0	0.0	3.3	0.0	0.0
	C2	0.0	6.7	0.0	0.0	0.0	83.3	0.0	0.0	10.0	0.0	0.0
	C3	0.0	3.3	0.0	0.0	0.0	93.3	0.0	3.3	0.0	0.0	0.0
	C4	0.0	0.0	3.3	0.0	0.0	96.7	0.0	0.0	0.0	0.0	0.0
	C5	0.0	3.3	0.0	0.0	0.0	96.7	0.0	0.0	0.0	0.0	0.0
Pinus	C1	0.0	0.0	0.0	0.0	0.0	0.0	97.5	0.0	0.0	2.5	0.0
	C2	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	C3	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	C4	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	C5	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Poaceae	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C3	0.0	0.0	0.0	2.5	0.0	0.0	0.0	97.5	0.0	0.0	0.0
	C4	0.0	0.0	0.0	2.5	0.0	0.0	0.0	97.5	0.0	0.0	0.0
	C5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Salix	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C2	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	97.5	0.0	0.0
	C3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C5	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	95.0	0.0	0.0
Lyc.	C1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C2	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	97.5	0.0
	C3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0

Fig. 5. Confusion matrix comparing predicted and true values for mixed-quality pollen images in percentages of all images per class. Five rows per predicted label across five cross-validation cycles are presented (C1–C5; Mahmood et al., 2023). The first split (purple) demonstrates the highest performing split. NPP = Non-pollen palynomorph.

example, F-measures were reported in Dunker et al. (2021), FTIR spectra match values in Pappas et al. (2003) and correlation coefficients in Mitsumoto et al. (2010) that cannot directly be compared to our recall values. In light of what was discussed, our object detection model shows promise in detecting the majority of pollen grains for most of the selected taxa from Late Glacial samples of temperate Europe.

All our five cross-validation cycles showed accuracy >95 %. The highest performing cross-validation cycle (C1) of our trained classification model with average accuracies of 100 % and 98.8 % for good quality images and mixed quality images, respectively, is comparable with other recent classification models for pollen of Daoud et al. (2018) with 100 % accuracy, Gonçalves et al. (2016) with 64 %, Lagerstrom et al. (2015) with 89 %, Marcos et al. (2015) with 95 %, Theuerkauf et al., 2023 (2023) with 93 % and Sevillano and Aznarte (2018) with 97

%. Accuracy values for classification between the different studies vary due to taxonomic resolution and number of pollen taxa (Olsson et al., 2021), as well as the state-of-the-art machine learning algorithms available at the time of previous studies (e.g., Lagerstrom et al., 2015; Li, 2004; Li; Sevillano and Aznarte, 2018).

With the exception of Theuerkauf et al., 2023, these studies focused on modern pollen samples and therefore, they did not encounter some of the complicating factors of fossil pollen assemblages (e.g., obscuring non-pollen debris) that our results were affected by. Overall, our classification model therefore compares favorably with these earlier studies. The trained model miss-classified a few images of poorer quality, similar to Bourel et al. (2020), who also reported lower accuracy rates on images of fossil grains and damaged pollen. The lower performance on images of Amaranthaceae (average error rate = 5.1 %) and

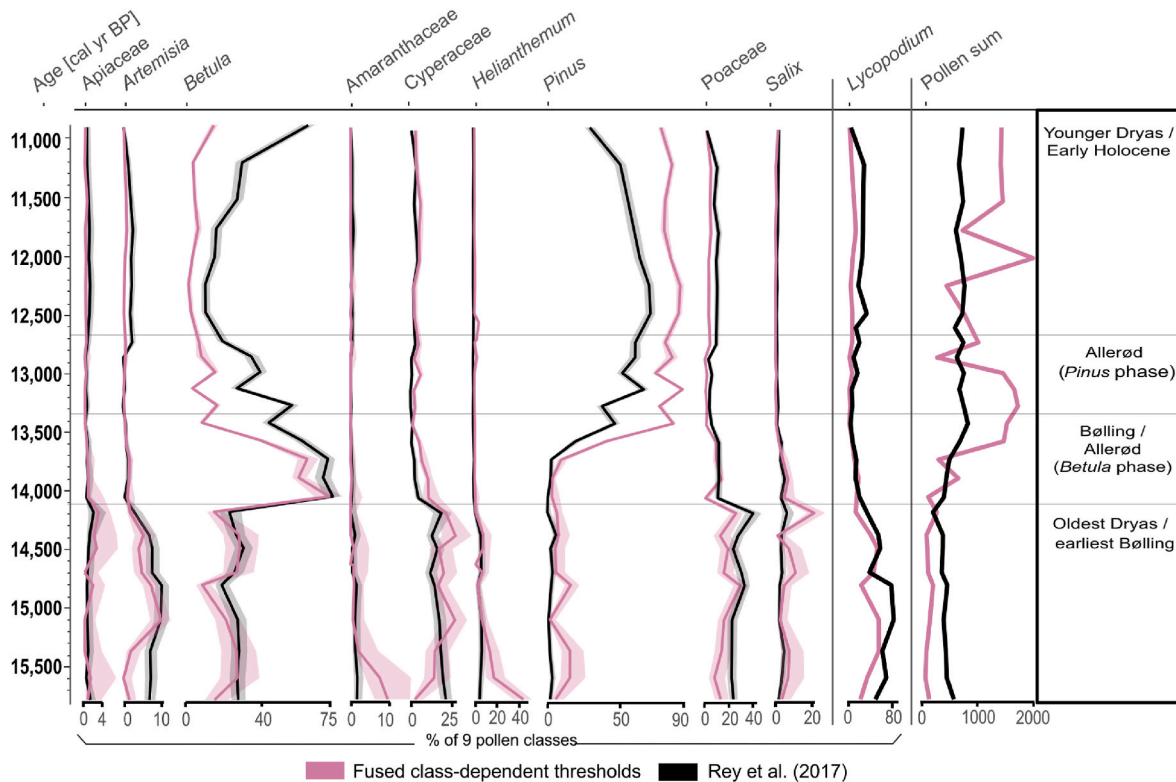


Fig. 6. Percentages based on automated and manual counts of nine pollen types and added *Lycopodium* spores (marker) in the Late Glacial sediments from the Burgäschisee record. Predicted percentages by the highest performing output configuration (fused class-dependent thresholds on classification and object detection scores; shown in purple) and manual counts (black curves; Rey et al., 2017). Pollen classes and *Lycopodium* are presented as percentages of the sum of all nine pollen classes. *Lycopodium* was excluded from the pollen sum. Transparent shading (purple and gray) indicates 95 % confidence interval for each automated and manually counted pollen type, respectively (Maher, 1972; Maher et al., 2012). The chronology follows Rey et al. (2017).

Helianthemum (average error rate = 7.3 %) can be explained by the lower amount of original training images available, highlighting some difficulties to establish training datasets for fossil pollen where some classes are underrepresented, and image recovery is challenging. The model had lower correct classification rates for the *Betula* class (97.5 % correctly classified), which probably resulted from insufficient image resolution to show characteristics that are critical for class affiliation. Furthermore, we simplified the *Betula* class by merging tree-type *Betula* with shrub-type *Betula* (Clegg et al., 2005) that can be distinguished by expert palynologists (e.g., Bos et al., 2007; Krüger and Damrath, 2020; Rey et al., 2017; Wick et al., 2024) due to the insufficient image resolution of the Zeiss AxioScan Z.1 (Carl Zeiss) scanning device used in this study. A separation of these types may lead to improved performance when identifying *Betula* class pollen in future studies.

4.2. Pipeline accuracy to develop recognition of fossil pollen

The final pipeline proposed here captures major Late Glacial pollen type shifts in the Burgäschisee record, including a phase with abundant herbaceous pollen in the oldest tundra samples from the Oldest Dryas, the high abundances in *Betula* pollen from the Bølling to the early Allerød, and the *Pinus*-dominated samples from the late Allerød, Younger Dryas, and Early Holocene. The high correlation between PCA axes 1 and 2 calculated between automated and manual pollen counts clearly indicates that the pollen stratigraphy captured long-term shifts in fossil pollen assemblages as deriving from vegetation changes. This was also confirmed by the high correlation between the PCA analyses indicated by the Procrustes rotation analysis. Our study, albeit at a selected taxonomic resolution, therefore, provides a first “proof of principle” that pollen stratigraphies comparable with manual counting may be developed in the future based on automated pollen counting procedures also

from complex fossil lake sediment samples and not only modern pollen or reference samples.

To the best of our knowledge, this is the first application of a fully automated pollen identification pipeline to develop stratigraphic data from fossil pollen samples (Bourel et al., 2020; Holt et al., 2011; Maffezzoli et al., 2023). Therefore, comparison is only possible with pipelines developed for the analysis of modern pollen assemblages or reference samples (e.g., Olsson et al., 2021; Oteros et al., 2020). The classification model in Olsson et al. (2021) was trained on reference material collected from 83 different plant taxa. By partitioning these taxa into 28 pollen types, the highest performing model achieved accuracy values of 79 %. Similarly, the state-of-the-art integrated system for automated pollen analysis, the Classifynder (Holt et al., 2011), uses 43 automatically calculated features that in more recent studies achieved average accuracies of 98 % (Sevillano et al., 2020). However, the Classifynder has, to our knowledge, not been successfully applied to fossil samples due to the large amount of organic non-pollen debris, the preservation conditions, and diversity of fossil pollen types in lake sediments (Holt et al., 2011). Although our pipeline captures trends caused by vegetation shifts, absolute differences in pollen percentages compared to manual counts are still relatively high for some taxa and individual samples. This can be expected for a first application of a newly developed automated identification approach to samples containing the full range of pollen types, NPPs, and other organic debris preserved in lake sediments.

Differences between the automated and manual counts can be attributed to systematic errors in image acquisition (i.e., pollen not on focal plane), errors in object detection (i.e., the system on average misses 11 % of pollen grains, also due to image quality), low classification confidence, and the three-dimensional geometry of pollen grains that is not accurately captured by a two-dimensional image. Holt et al. (2011)

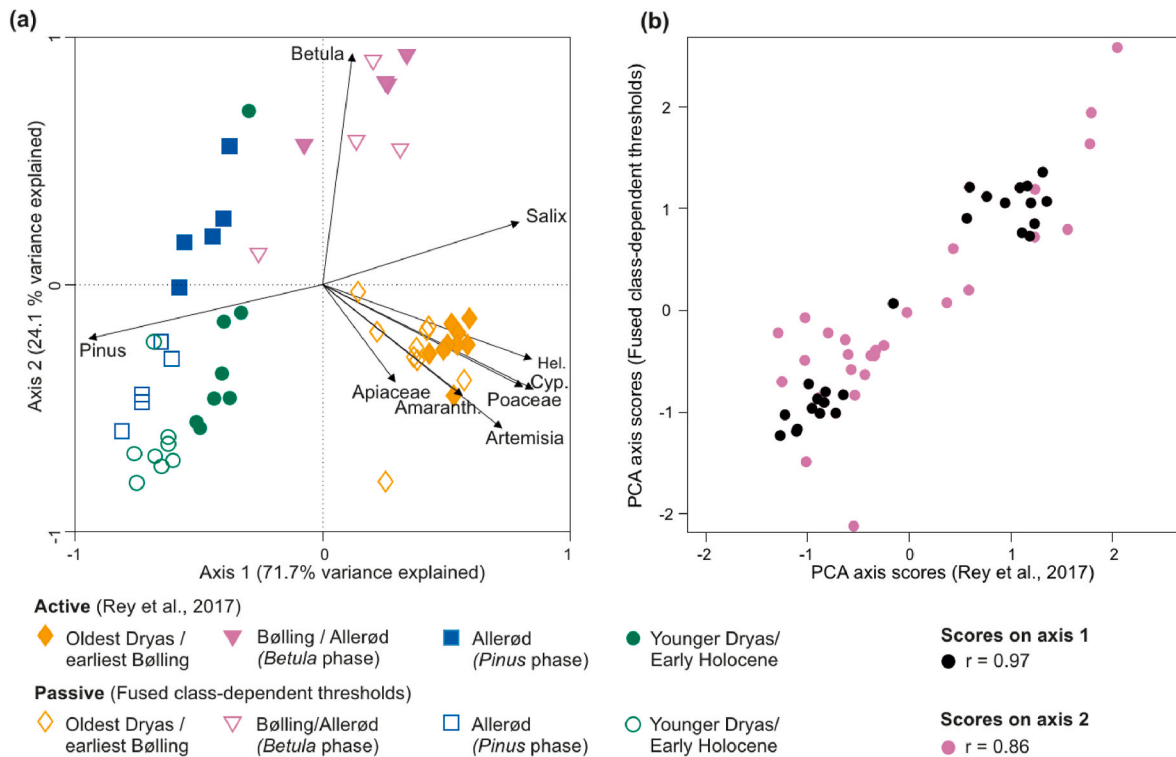


Fig. 7. Principal Component Analyses (PCA) for automated (Fused class-dependent thresholds) and manually (Rey et al., 2017) counted pollen percentages of nine pollen classes in the Late Glacial pollen record from Burgäschisee. (A) Biplot of sample and species scores calculated with the manual counts of Rey et al. (2017). Samples based on automated identification are passively plotted in the PCA. Distances between samples approximate sample dissimilarity. (B) Comparison of PCA axis scores for two separate PCAs, the first based on manually identified pollen in the Burgäschisee record, the second based on automated pollen identification.

discuss that pollen counts of different palynologists may also vary to some extent due to differences in pollen counting strategies, the amount of pollen counted and vagaries in identification. Although these differences may be much smaller than systematic errors during automated image analysis using our approach, systematic errors of automated pipelines are consistent such that automatically-generated records will be highly comparable with each other (Marcos et al., 2015). If these errors would be further reduced due to progresses in automated recognition, automatically generated pollen records may in the future deliver the high palynological quality needed to establish correct biostratigraphies and vegetation history reconstructions.

Currently, our pipeline is not able to generate a pollen record with the taxonomic resolution of a manual record. Specifically, we focused on few important pollen classes in Late Glacial sequences and merged similar pollen types of the same family or genus to one class. Rare taxa with single occurrences were not implemented as a separate class due to the lack of training data. Due to the limited image resolution, the pipeline could not be trained for *Juniperus*, an important taxon of Late Glacial sequences in Europe (Lotter, 1999; Ammann et al., 2013; Duprat-Oualid et al., 2017; Rey et al., 2017; Wick et al., 2024), because crucial information to accurately identify *Juniperus* characteristics is lost in images of lower pixel resolution.

Several areas for possible technical improvements emerged from our results. First, the image resolution used for the classifier in this study was smaller than in previous studies on modern pollen assemblages (e.g., Lagerstrom et al., 2015; Polling, 2021; Sevillano et al., 2020; Valiente et al., 2023). Hence, it can be expected that higher resolution images would improve object detection and classification. Second, the number of training images for training the object detection and classification models were limited to 35 images per class in this study, as a result of visually annotating ~200,000 images. The number of images was comparatively low, albeit in a similar range as previous studies (Astolfi et al., 2020; Li, 2004; Sevillano and Aznarte, 2018) since it was difficult

to find sufficient high-quality images. Major efforts for expanding this training image dataset would also improve the performance of the pipeline. Third, it can be expected that overall fossil pollen grains are presently more challenging for the automated pipeline to identify than for manual analysts that are used to and can account for folded, obscured or strangely oriented pollen grains during identification. Improvements of performance would be possible by treating poorly oriented pollen from a particular pollen taxon as separate classes and merging these for the final compilation of the results. For example, in this study we already treated pollen grains of *Artemisia*, *Poaceae*, and *Betula* in polar and equatorial view as separate pollen classes. Fourth, a higher number of pollen counts could be relatively easily implemented by imaging several slides. This would ensure a higher number of well-placed pollen grains per sample and taxon, and therefore increase the number of good quality images.

In summary, several advancements could potentially improve the performance of our approach in future applications. These include increased resolution of images during pollen imaging, an expansion of the number and quality of images used to train the image detection and classification procedures, and treating different orientations, preservation states or quality of images as separate classes during training before combining them into categories. In addition, considering the rapid advances in the field in recent years, major advances in CNN-based image detection and classification models are expected in the near future, which will further increase the accuracy and precision of automated identifications.

Our fossil image dataset consisted of less pollen classes compared with other recent studies on modern pollen that used over 40 different classes (Oteros et al., 2020; Olsson et al., 2021; Sevillano et al., 2020), and considerably fewer classes than count data from manual analysts that often separate hundreds of different pollen and other NPP types in individual records (e.g., Rey et al., 2017, 2022; Wick et al., 2024). Expanding the number of pollen classes in training datasets for Late

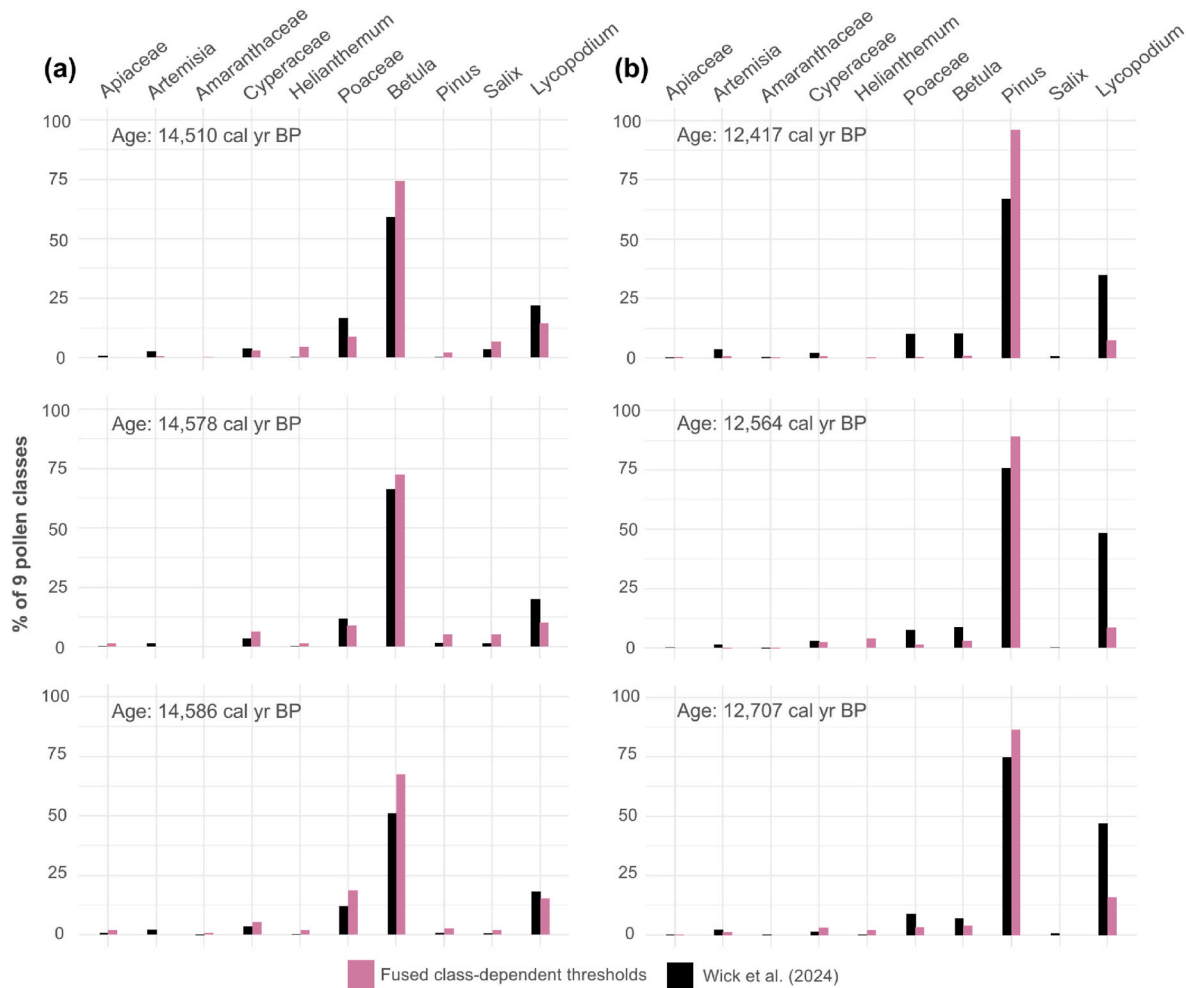


Fig. 8. Percentages of pollen classes in Hüttwilersee samples during the Bølling (A) and Younger Dryas (B). Pollen percentages based on automated pollen counts (purple; configuration with fused class-dependent thresholds) and comparison with manual counts (black; Wick et al., 2024). Pollen classes and *Lycopodium* are shown as percentages of the sum of all nine pollen classes. *Lycopodium* markers were excluded from the pollen sum.

Glacial sequences and expanding our approach, or similar approaches (Bourel et al., 2020), to Late Quaternary sequences from different time intervals (e.g., Early Holocene to Mid Holocene forest types, Late Holocene agricultural landscapes) and biogeographic regions (e.g., steppe or arctic followed by temperate, subtropical, and tropical) are therefore next challenging steps, considering that pollen diversity increases markedly towards the present and in warmer biomes (e.g., Brugger et al., 2019b, 2019c; Rinta et al., 2016; Schwörer et al., 2022).

Our study shows that building suitable training data for classes by extracting images from fossil pollen assemblages is extremely time consuming and such efforts increase further for rare pollen types as each pollen image has to be manually checked and assigned to the respective class for quality control. For example, to obtain 35 training images per each of the ten classes in this study, ~200,000 images were manually checked for image quality and manually annotated with the respective name of the pollen class. Expanding this approach to more classes, and particularly also including pollen in different orientations per class, as suggested above, would massively expand this workload. Therefore, we argue that before expanding the overall scope of such efforts to a large range of pollen classes, applications to a more limited number of taxa and specific time intervals, as provided in this study, may be helpful to optimize future efforts and assign resources to refining the most relevant steps of image detection and classification. For example, similar studies covering the (rather species-poor) Late Glacial, or other time intervals with clearly defined vegetation changes, may help to assess, which

optimization steps may lead to a significant increase in detection and classification quality or cause unfruitful efforts (e.g., increase in scanning resolution, expansion of training images to poorly preserved or unusual specimens, separate classifications of pollen grains of particular taxa in different orientations and preservation).

5. Conclusions

We present a first study and “proof of principle” that demonstrates the application of automated pollen detection and classification techniques to develop stratigraphies from fossil pollen samples that include the full range of organic matter in lacustrine environments. Our proposed pipeline for detecting and classifying pollen-like objects recreated the pollen signals of a typical European Late Glacial sequence despite the inherent challenges presented by fossil pollen assemblages compared with modern pollen samples. Our automatically generated pollen record using this pipeline is still limited and does not yet reach the taxonomic resolution or the accuracy of manually counted pollen records. However, it clearly shows main vegetation trends in the examined Burgäschisee record and we also demonstrate success in applying the developed approach to a Late Glacial sequence from a different lake (Hüttwilersee), thereby revealing the transferability of the approach to independent localities and settings. Several developments and expansions can be expected to increase the performance and applicability of our approach, or similar approaches. Advancements may include further

improvements and developments of CNN-based automated image detection and classification procedures, increases in image resolution and quality, and expansion of our method to a higher number of pollen classes with images of different quality and orientation. At the same time our results and the application to Late Glacial sediments indicate that the developed pipeline can already now be used to gain an initial overview of the abundance of major pollen types in a new sediment sequence from similar (i.e., European Late Glacial) settings before expert analysts use time and resources for manually counting a detailed record. New models and algorithms are constantly being developed across several scientific disciplines, some specifically for pollen analysis (Mahmood et al., 2023), that may minimize computing consumption and processing time. These efforts will provide new tools for further incorporating automated pipelines into classical paleoecology in the near future.

CRediT authorship contribution statement

Robin von Allmen: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. **Sandra O. Brugger:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Visualization, Writing – original draft. **Kai D. Schleicher:** Conceptualization, Investigation, Visualization, Writing – review & editing. **Fabian Rey:** Validation, Writing – review & editing. **Erika Gobet:** Validation, Writing – review & editing. **Colin J. Courtney Mustaphi:** Conceptualization, Validation, Writing – review & editing. **Willy Tinner:** Validation, Writing – review & editing. **Oliver Heiri:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Writing – original draft.

Declaration of competing interest

We declare no competing interests for the submitted manuscript and the related dataset. Furthermore, we confirm that all co-authors have made substantial contributions and approved the final version of the manuscript.

Data availability

Code and training images are available on GitHub and Zenodo, respectively and active links are provided in the manuscript file.

Acknowledgements

We thank Lucia Wick for discussions about pollen identification criteria. Pollen sample acquisition from Hüttwilersee was supported by the Walter Enggist-Fonds and Kantonsarchäologie Kanton Thurgau via the KUMIT project. This study received funding by the Swiss National Science Foundation (SNSF-Ambizione project grant PZ00P2_208687). We are grateful to H. John B. Birks for valuable comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.quascirev.2024.108521>.

References

- Abadi, M., et al., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265–283. <https://doi.org/10.48550/arXiv.1603.04467>. Preprint at arXiv.
- Adaïmé, M.É., Kong, S., Punyasena, S.W., 2023. Deep Learning Approaches to the Phylogenetic Placement of Extinct Pollen Morphotypes. *bioRxiv*, p. 7. <https://doi.org/10.1101/2023.07.09.545296>, 2023.
- Ammann, B., van Leeuwen, J.F., van der Knaap, W.O., Lischke, H., Heiri, O., Tinner, W., 2013. Vegetation responses to rapid warming and to minor climatic fluctuations during the Late-Glacial Interstadial (GI-1) at Gerzensee (Switzerland). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 391, 40–59. <https://doi.org/10.1016/j.palaeo.2012.07.010>.
- Astolfi, G., et al., 2020. POLLEN73S: an image dataset for pollen grains classification. *Ecol. Inf.* 60, 101165 <https://doi.org/10.1016/j.ecoinf.2020.101165>, 2020.
- Beug, H.J., 2004. *Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete*. Friedrich Pfeil, München.
- Birks, H.J.B., et al., 2012. *Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*, fifth ed. Springer, Dordrecht.
- Bos, J.A., van Geel, B., van der Plicht, J., Bohncke, S.J., 2007. Preboreal climate oscillations in Europe: Wiggle-match dating and synthesis of Dutch high-resolution multi-proxy records. *Quat. Sci. Rev.* 26 (15–16), 1927–1950. <https://doi.org/10.1016/j.quascirev.2006.09.012>.
- Bourel, B., et al., 2020. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. *Comput. Geosci.* 140, 104498 <https://doi.org/10.1016/j.cageo.2020.104498>.
- Brechbühl, S., et al., 2023. Vegetation dynamics and land-use change at the Neolithic lakeshore settlement site of Ploča Mičov Grad, lake Ohrid, North Macedonia. *Veg. Hist. Archaeobotany* 1–21. <https://doi.org/10.1007/s00334-023-00931-3>.
- Brugger, S.O., et al., 2019a. Why loss matters: Reply to the comments of Festi and others on ‘A quantitative comparison of microfossil extraction methods from ice cores’ by Brugger and others (2018). *J. Glaciol.* 65 (253), 867–868. <https://doi.org/10.1017/jog.2019.61>.
- Brugger, S.O., et al., 2019b. Tropical Andean glacier reveals colonial legacy in modern mountain ecosystems. *Quat. Sci. Rev.* 220, 1–13. <https://doi.org/10.1016/j.quascirev.2019.06.032>.
- Camuera, J., et al., 2019. Vegetation and climate changes during the last two glacial-interglacial cycles in the western Mediterranean: a new long pollen record from Padul (southern Iberian Peninsula). *Quat. Sci. Rev.* 205, 86–105. <https://doi.org/10.1016/j.quascirev.2018.12.013>.
- Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. <https://doi.org/10.48550/arXiv.1610.02357>, 6569–6578.
- Chollet, F., 2020. Grad-CAM class activation visualization. https://keras.io/examples/vison/grad_cam/.
- Clegg, B.F., Tinner, W., Gavin, D.G., Hu, F.S., 2005. Morphological differentiation of Betula (birch) pollen in northwest North America and its palaeoecological application. *Holocene* 15 (2), 229–237. <https://doi.org/10.1191/0959683605hl788rp>.
- Clow, M., 2018. Visual Studio code. In: *Angular 5 Projects: Learn to Build Single Page Web Applications Using 70+ Projects*, vols. 57–68. Apress, Berkeley. https://doi.org/10.1007/978-1-4842-3279-8_5.
- Daoud, A., Ribeiro, E., Bush, M., 2018. Sequential recognition of pollen grain Z-Stacks by combining CNN and RNN. In: *The Thirty-First International Flairs Conference*, 2018. <https://www.aaii.org/ocs/index.php/FLAIRS/FLAIRS18/paper/view/17642>.
- Daunys, G., Šukienė, L., Vaitkevičius, L., Valiulis, G., Sofiev, M., Šaulienė, I., 2022. Comparison of computer vision models in application to pollen classification using light scattering. *Aerobiologia* 1. <https://doi.org/10.1007/s10453-022-09769-0>. –13.
- Dell’Anna, R., et al., 2009. Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. *Anal. Bioanal. Chem.* 394 (5), 1443–1452. <https://doi.org/10.1007/s00216-009-2794-9>.
- Dendiev, A.M., Serieyssel, K., Dietre, B., Cubizolle, H., Quiquerez, A., Haas, J.N., 2022. Late-glacial and early holocene environmental changes affecting the shallow lake basin of la narce du béage (ardèche, massif central, France). *Quat. Int.* 636, 52–68. <https://doi.org/10.1016/j.quaint.2019.09.014>.
- D’Orazio, M., et al., 2022. Electro-optical classification of pollen grains via microfluidics and machine learning. *IEEE Trans. Biomed. Eng.* 69 (2), 921–931. <https://doi.org/10.1109/TBME.2021.3109384>.
- Duan, K., et al., 2019. CenterNet: Keypoint Triplets for Object Detection. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea. <https://doi.org/10.1109/ICCV.2019.00667>, 6568–6577.
- Dunker, S., et al., 2021. Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytol.* 229 (1), 593–606. <https://doi.org/10.1111/nph.16882>.
- Duprat-Qualid, F., Rius, D., Bégeot, C., Magny, M., Millet, L., Wulf, S., Appelt, O., 2017. Vegetation response to abrupt climate changes in Western Europe from 45 to 14.7 k cal a BP: the Bergsee lacustrine record (Black Forest, Germany). *J. Quat. Sci.* 32 (7), 1008–1021. <https://doi.org/10.1002/jqs.2972>.
- El Azari, H., et al., 2023. A laboratory evaluation of the new automated pollen Sensor beonose: pollen discrimination using machine learning techniques. *Sensors* 23 (6), 2964. <https://doi.org/10.3390/s23062964>.
- Felde, V.A., Birks, H.H., 2019. Using species attributes to characterize late-glacial and early-Holocene environments at Kråkenes, western Norway. *J. Veg. Sci.* 30 (6), 1228–1238. <https://doi.org/10.1111/jvs.12804>.
- Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., Li, B., 2020. Axiom-based grad-cam: towards accurate visualization and explanation of cnns. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2008.02312>.
- Galka, M., et al., 2022. Insight into the factors of mountain bog and forest development in the Schwarzwald Mts.: implications for ecological restoration. *Ecol. Indic.* 140, 109039 <https://doi.org/10.1016/j.ecolind.2022.109039>.
- Gallardo-Caballero, R., García-Orellana, C.J., García-Manso, A., González-Velasco, H.M., Tormo-Molina, R., Macías-Macias, M., 2019. Precise pollen grain detection in bright field microscopy using deep learning techniques. *Sensors* 19 (16), 3583. <https://doi.org/10.3390/s19163583>.
- Githumbi, E.N., et al., 2018. Pollen, people and place: multidisciplinary perspectives on ecosystem change at Amboseli, Kenya. *Frontiers Earth Sci* 5, 113. <https://doi.org/10.3389/feart.2017.00113>.

- Gonçalves, A.B., Souza, J.S., Silva, G.G.D., Cereda, M.P., Pott, A., Naka, M.H., Pistori, H., 2016. Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. *PLoS One* 11 (6), e0157044. <https://doi.org/10.1371/journal.pone.0157044>.
- He, H., Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* 21, 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Holt, K.A., Bebbington, M.S., 2014. Separating morphologically similar pollen types using basic shape features from digital images: a preliminary study. *Appl. Plant Sci.* 2 (8), 1400032. <https://doi.org/10.3732/apps.1400032>.
- Holt, K.A., Bennett, K.D., 2014. Principles and methods for automated palynology. *New Phytol.* 203 (3), 735–742. <https://doi.org/10.1111/nph.12848>.
- Holt, K., et al., 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Rev. Palaeobot. Palynol.* 167 (3–4), 175–183. <https://doi.org/10.1016/j.revpalbo.2011.08.006>.
- Ivleva, N.P., et al., 2005. Characterization and discrimination of pollen by Raman microscopy. *Anal. Bioanal. Chem.* 381 (1), 261–267. <https://doi.org/10.1007/s00216-004-2942-1>.
- Japkowicz, N., Stephen, S., 2002. The class imbalance problem: a systematic study. *Intell. Data Anal.* 6 (5), 429–449. <https://doi.org/10.3233/IDA-2002-6504>.
- Jermain, C.L., et al., 2016. GPU-accelerated micromagnetic simulations using cloud computing. *J. Magn. Magn.* 401, 320–322. <https://doi.org/10.1016/j.jmmm.2015.10.054>.
- Johnsrud, S., et al., 2013. Semi-automated segmentation of pollen grains in microscopic images: a tool for three imaging modes. *Grana* 52 (3), 181–191. <https://doi.org/10.1080/00173134.2013.768291>.
- Kamal, I.M., Salah Abd-elhafeez, M., Farghal, A., 2023. Camera-based navigation system for blind and visually impaired people. *S. E. J.* 3 (1), 1–13. <https://doi.org/10.21608/sej.2022.155927.1018>.
- Khanzhina, N., Putin, E., Filchenkov, A., Zamyatina, E., 2018. Pollen grain recognition using convolutional neural network. In: *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. ESANN 2018). Bruges, Belgium, 409–414. <https://www.esann.org/sites/default/files/proceedings/legacy/es2018-199.pdf>.
- Kim, S., Wimmer, H., Kim, J., 2022. Analysis of Deep Learning Libraries: Keras, PyTorch, and MXnet. 2022. *IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 54–62. <https://doi.org/10.1109/SERA54885.2022.9806734>.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint* 1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>.
- Krüger, S., Dammrath, M., 2020. In search of the Bolling-Oscillation: a new high resolution pollen record from the locus classicus Lake Bolling, Denmark. *Veg. Hist. Archaeobotany* 29, 189–211. <https://doi.org/10.1007/s00334-019-00736-3>.
- Kubera, E., Kubik-Komar, A., Kurasinski, P., Piotrowska-Weryszko, K., Skrzypiec, M., 2022. Detection and recognition of pollen grains in multilabel microscopic images. *Sensors* 22 (7), 2690. <https://doi.org/10.3390/s22072690>.
- Lagerstrom, R., et al., 2015. Pollen image classification using the classifyfynder system: algorithm comparison and a case study on New Zealand honey. In: Sun, C., Bednarz, T., Pham, T.D., Vallotton, P., Wang, D. (Eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, 207–226. Springer, Cham. https://doi.org/10.1007/978-3-319-10984-8_12.
- Lang, G., et al., 2023. *Quaternary Vegetation Dynamics of Europe*. Bern, Haupt.
- Li, P., 2004. Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. *J. Quat. Sci.* 19 (8), 755–762. <https://doi.org/10.1002/jqs.874>.
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: common objects in context. In: *Computer Vision—ECCV 2014: 13th European Conference, vol. 13*. Springer International Publishing, Zurich, Switzerland, pp. 740–755. September 6–12, 2014, *Proceedings, Part V*.
- Linkert, M., Rueden, C.T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., et al., 2010. Metadata matters: access to image data in the real world. *J. Cell Biol.* 189, 777–782. <https://doi.org/10.1083/jcb.201004104>.
- Lotter, A.F., 1999. Late-glacial and Holocene vegetation history and dynamics as shown by pollen and plant macrofossil analyses in annually laminated sediments from Soppensee, central Switzerland. *Veg. Hist. Archaeobotany* 8, 165–184. <https://doi.org/10.1007/BF02342718>.
- Maffezzoli, N., et al., 2023. Detection of ice core particles via deep neural networks. *Cryosphere* 17 (2), 539–565. <https://doi.org/10.5194/tc-17-539-2023>.
- Maher Jr., L.J., 1972. Nomograms for computing 0.95 confidence limits of pollen data. *Rev. Palaeobot. Palynol.* 13 (2), 85–93. [https://doi.org/10.1016/0034-6667\(72\)90038-3](https://doi.org/10.1016/0034-6667(72)90038-3).
- Maher, L.J., Heiri, O., Lotter, A.F., 2012. Assessment of uncertainties associated with palaeolimnological laboratory methods and microfossil analysis. In: Birks, H., Lotter, A., Juggins, S., Smol, J. (Eds.), *Tracking Environmental Change Using Lake Sediments, Developments in Paleoenvironmental Research*, vol. 5. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2745-8_6.
- Mahmood, T., et al., 2023. Artificial intelligence-based classification of pollen grains using attention-guided pollen features aggregation network. *J. King Saud Univ. - Comput. Inf.* 35, 740–756. <https://doi.org/10.1016/j.jksuci.2023.01.013>.
- Marcos, J.V., et al., 2015. Automated pollen identification using microscopic imaging and texture analysis. *Micron* 68, 36–46. <https://doi.org/10.1016/j.micron.2014.09.002>.
- Mathai-Davis, S., Trudeau, E., 2019. Bitcoin systematic trading algorithms in the cloud: challenges and opportunities. 2019 *IEEE Cloud Summit* 25–30. <https://doi.org/10.1109/CloudSummit47114.2019.00011>.
- Matsuo, M., et al., 2022. Automated focal plane merging from a stack of gonioscopic photographs using a focus-stacking algorithm. *Trans. Vis. Sci. Tech.* 11 (4) <https://doi.org/10.1167/tvst.11.4.22>, 22–22.
- Milošević, D., et al., 2020. Application of deep learning in aquatic bioassessment: towards automated identification of non-biting midges. *Sci. Total Environ.* 711, 135160. <https://doi.org/10.1016/j.scitotenv.2019.135160>.
- Mitsumoto, K., et al., 2010. Development of a novel real-time pollen-sorting counter using species-specific pollen autofluorescence. *Aerobiologia* 26 (2), 99–111. <https://doi.org/10.1007/s10453-009-9147-1>.
- Oksanen, J., et al., 2022. Vegan: community ecology package, 2.6-4. <https://CRAN.R-project.org/package=vegan>.
- Olsson, O., Karlsson, M., Persson, A.S., Smith, H.G., Varadarajan, V., Yourstone, J., Stjernman, M., 2021. Efficient, automated and robust pollen analysis using deep learning. *Methods Ecol. Evol.* 12 (5), 850–862. <https://doi.org/10.1111/2041-210X.13575>.
- Oteros, J., et al., 2015. Automatic and online pollen monitoring. *Int. Arch. Allergy Immunol.* 167 (3), 158–166. <https://doi.org/10.1159/000436968>.
- Oteros, J., et al., 2020. An operational robotic pollen monitoring network based on automatic image recognition. *Environ. Res.* 191, 110031. <https://doi.org/10.1016/j.envres.2020.110031>.
- Pan, S.J., Yang, Q., 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22 (10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>.
- Pappas, C.S., Tarantilis, P.A., Harizanis, P.C., Polissiou, M.G., 2003. New method for pollen identification by FT-IR spectroscopy. *Appl. Spectrosc.* 57 (1), 23–27.
- Pardasani, R., Awasthi, N., 2020. Classification of 12 lead ECG signal using 1D-convolutional neural network with class dependent threshold. *Computing in Cardiology IEEE 1*. <https://doi.org/10.22489/CinC.2020.277>, 4.
- Pennington, W., 1996. Limnic sediments and the taphonomy of Lateglacial pollen assemblages. *Quat. Sci. Rev.* 15 (5–6), 501–520. [https://doi.org/10.1016/0277-3791\(96\)00019-4](https://doi.org/10.1016/0277-3791(96)00019-4).
- Peres-Neto, P.R., Jackson, D.A., 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 129, 169–178. <https://doi.org/10.1007/s004420100720>.
- Polling, M., 2021. *The Hidden Biodiversity of Pollen*. Doctoral dissertation. University of Oslo, Norway. <https://repository.naturalis.nl/pub/800041/>.
- Rey, F., et al., 2017. Vegetational and agricultural dynamics at Burgäschisee (Swiss Plateau) recorded for 18,700 years by multi-proxy evidence from partly varved sediments. *Veg. Hist. Archaeobotany* 26, 571–586. <https://doi.org/10.1007/s00334-017-0635-x>.
- Rey, F., et al., 2019. Causes and mechanisms of synchronous succession trajectories in primeval Central European mixed Fagus sylvatica forests. *J. Ecol.* 107 (3), 1392–1408. <https://doi.org/10.1111/1365-2745.13121>.
- Rey, F., et al., 2022. 14,500 years of vegetation and land use history in the upper continental montane zone at Lac de Champey (Valais, Switzerland). *Veg. Hist. Archaeobotany* 31, 377–393. <https://doi.org/10.1007/s00334-021-00859-6>.
- Rinta, P., et al., 2016. Land use affects carbon sources to the pelagic food web in a small boreal lake. *PLoS One* 11 (8), e0159900. <https://doi.org/10.1371/journal.pone.0159900>.
- Schwörer, C., et al., 2022. Holocene vegetation, fire and land use dynamics at Lake Svityaz, an agriculturally marginal site in northwestern Ukraine. *Veg. Hist. Archaeobotany* 31 (2), 155–170. <https://doi.org/10.1007/s00334-021-00844-z>.
- Sevillano, V., Aznarte, J.L., 2018. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. *PLoS One* 13 (6), e0201807. <https://doi.org/10.1371/journal.pone.0201807>.
- Sevillano, V., Holt, K., Aznarte, J.L., 2020. Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLoS One* 15 (6), e0229751. <https://doi.org/10.1371/journal.pone.0229751>.
- Simpson, G.L., Birks, H.J.B., 2012. *Statistical Learning in Palaeolimnology. Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*. Springer, Dordrecht, pp. 249–327. https://doi.org/10.1007/978-94-007-2745-8_9.
- Srivastava, S., et al., 2022. Sign Language recognition system using TensorFlow object detection API. In: *Advanced Network Technologies and Intelligent Computing*. Springer International Publishing, Cham, pp. 634–646. https://doi.org/10.1007/978-3-030-96040-7_48.
- Stillman, E.C., Flenley, J.R., 1996. The needs and prospects for automation in palynology. *Quat. Sci. Rev.* 15 (1), 1–5. [https://doi.org/10.1016/0277-3791\(95\)00076-3](https://doi.org/10.1016/0277-3791(95)00076-3).
- Subramanian, M., Shanmugavadivel, K., Nandhini, P.S., 2022. On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Comput. Appl.* 34 (16), 13951–13968. <https://doi.org/10.1007/s00521-022-07246-w>.
- Ter Braak, C.J.F., Smilauer, P., 1997–2019. *Software for Multivariate Data Exploration, Testing, and Summarization*. In: Canoco 5, Windows Release (5.12). Biometris, Wageningen.
- Theuerkauf, M., Siradze, N., Gillert, A., 2023. A trainable object finder, selector and identifier for pollen, spores and other things: A step towards automated pollen recognition in lake sediments. *Holocene*. <https://doi.org/10.1177/09596836231211876>.
- Torres, V., et al., 2013. Astronomical tuning of long pollen records reveals the dynamic history of montane biomes and lake levels in the tropical high Andes during the Quaternary. *Quat. Sci. Rev.* 63, 59–72. <https://doi.org/10.1016/j.quascirev.2012.11.004>.
- Tzatalin, D., 2023. *Labellmg-LabelImg* is a graphical image annotation tool. Github repository. <https://github.com/tzatalin/labelmg>.

- Valiente, J.M., Juan-Borrás, M., López-García, F., Escriche, I., 2023. Automatic pollen recognition using convolutional neural networks: the case of the main pollens present in Spanish citrus and rosemary honey. *J. Food Compos. Anal.* 105605 <https://doi.org/10.1016/j.jfca.2023.105605>.
- Wick, L., Rey, F., Heiri, O., 2024. Pollenanalyse und Pflanzenmakroreste. In: Geisser, H., Leuzinger, U. (Eds.), *Klima, Umwelt, Mensch im Thurgau (KUMIT) – hochaufgelöste Umweltrekonstruktionen der letzten 17'000 Jahre anhand von Sedimentkernen aus dem Bichelsee und Hüttwilersee, Kanton Thurgau (Schweiz). Mitteilungen der Thurgauischen Naturforschenden Gesellschaft (Band 72).* Frauenfeld.
- Yamazaki, A., et al., 2023. A simple method for measuring pollen germination rate using machine learning. *Plant Reprod.* <https://doi.org/10.1007/s00497-023-00472-9>.
- Zeng, H., et al., 2023. CNN-based EEG classification method for drug use detection. In: *Proceedings of the 2022 11th International Conference on Computing and Pattern Recognition (S. 418–423).* Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3581807.3581868>.